

IMDb Movie Reviews Sentiment Analysis

Anish Rao
20066423



Dataset & Project Objectives

Dataset

IMDb movie reviews
50,000 reviews

Pre Processing

Preprocessing steps included removing duplicates, HTML tags, and punctuation. Text was lowercased and stop-words were eliminated. These steps were crucial for preparing the data for effective sentiment analysis.

After data cleaning, 49,582 reviews were retained. The balanced nature of the dataset ensures unbiased analysis and reliable results.

Objectives

1

Identify if reviews are mostly positive or negative

2

Common words in positive vs. negative reviews

3

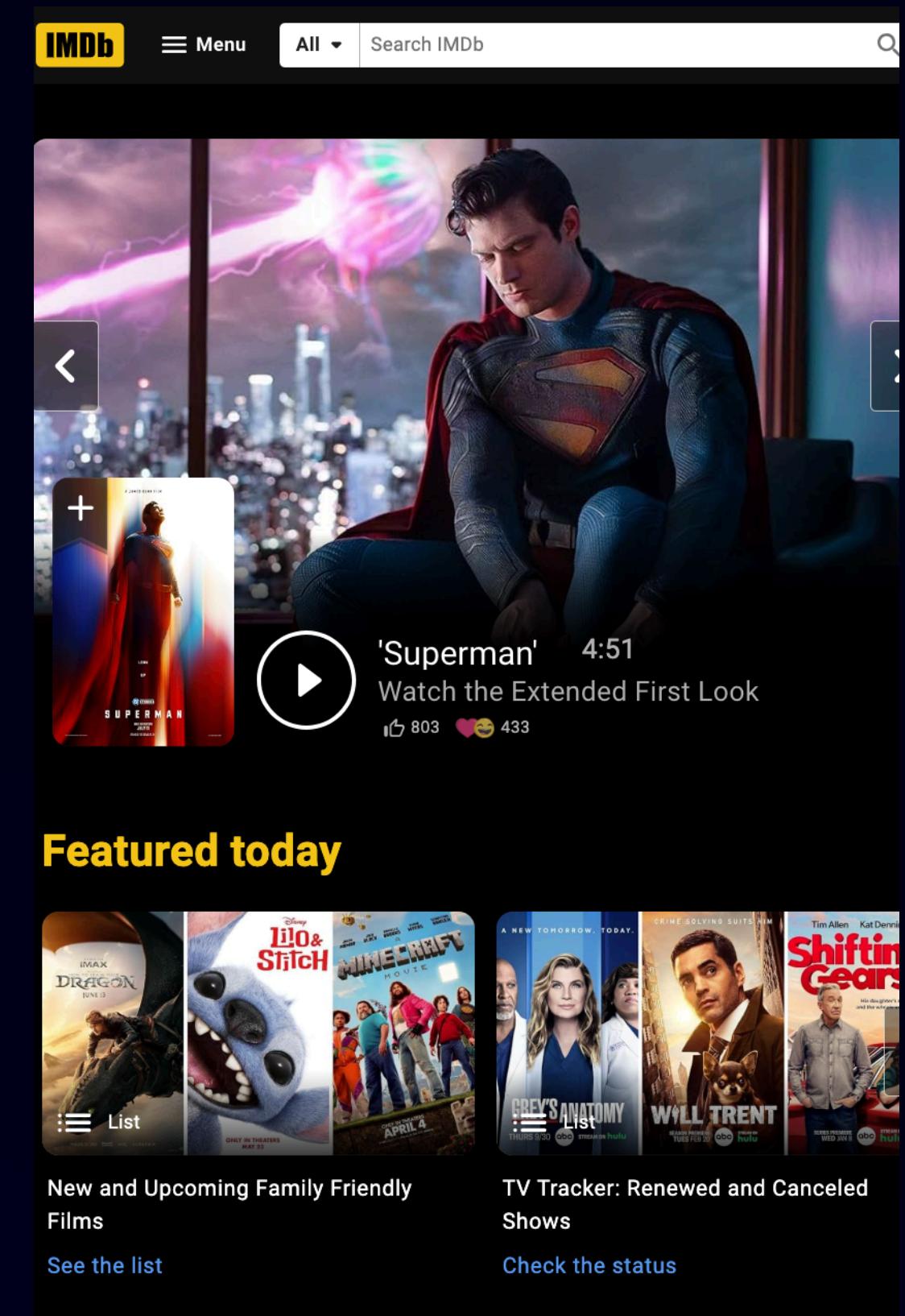
Relationship between review length and sentiment

4

Polarity and subjectivity (emotional vs. neutral opinions)

5

Patterns in extreme sentiment reviews





Key Results

50.2%

Positive Sentiment

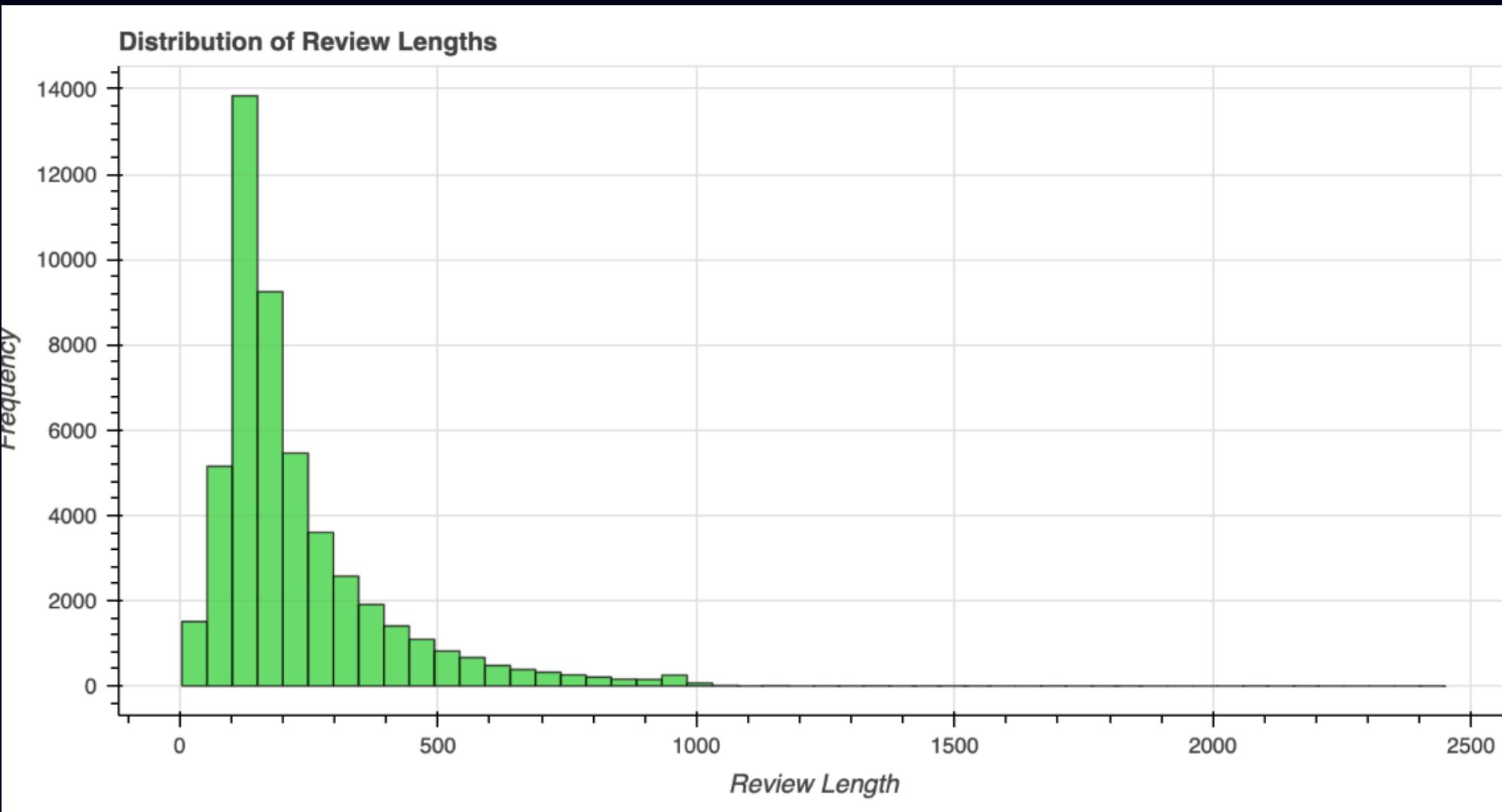


49.8%

Positive vocabulary includes words like "great," "fun," and "best". Negative vocabulary includes "boring," "bad," and "worst".

Common words suggests that tone and context significantly differ.

Review Length Analysis



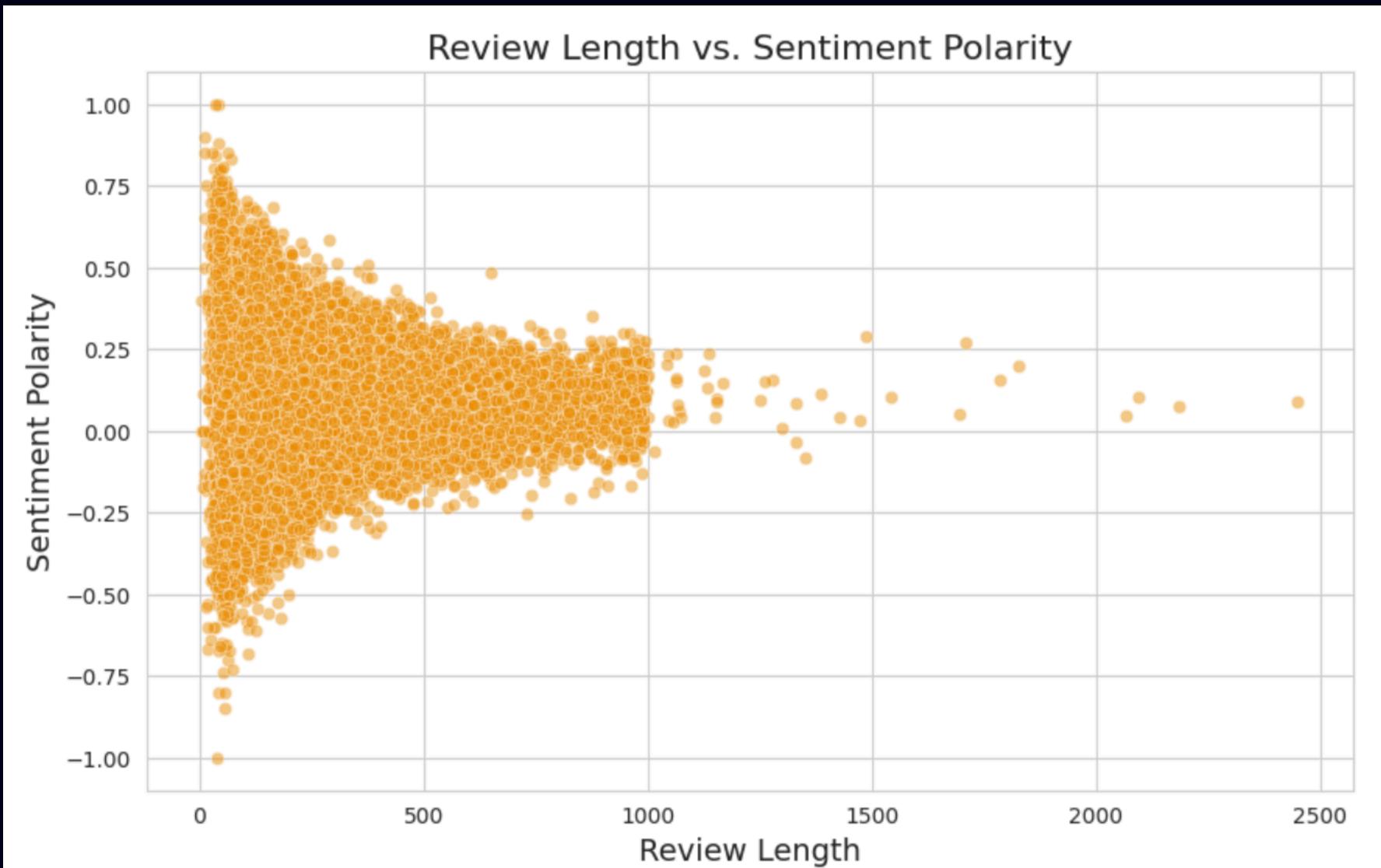
100-200

Review Length
Most reviews in words

Most reviews are short, with 46% reviews between 100–200 words. Very long reviews are rare with only 41 over 1,000 words.

Positive reviews show slightly more length variation, but overall, sentiment has little impact on review length.

Sentiment Intensity vs Review Length



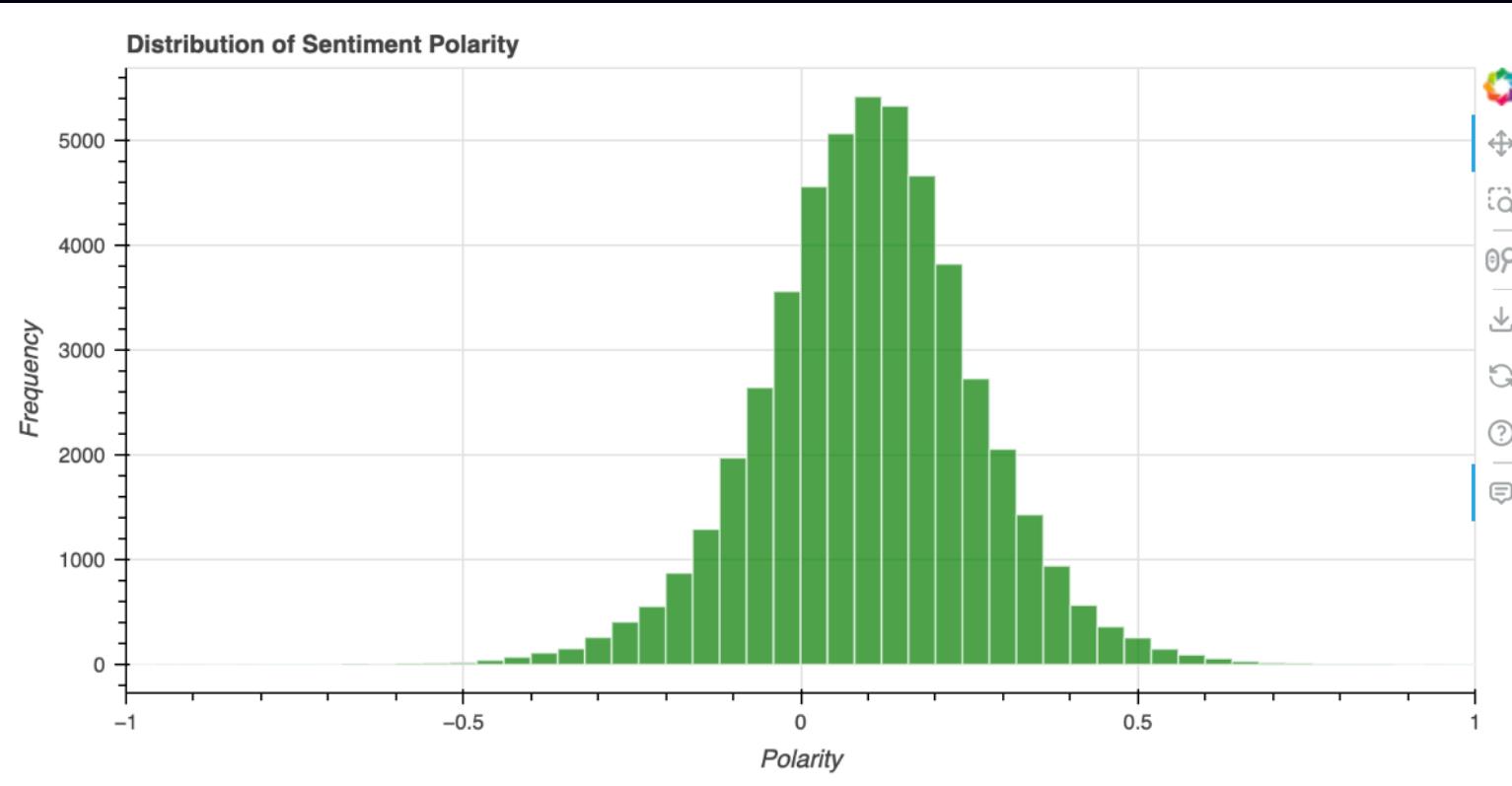
Weak negative correlation (-0.05) between review length and sentiment polarity

Short Reviews (< 200)

Emotionally stronger

Long Reviews (> 500)

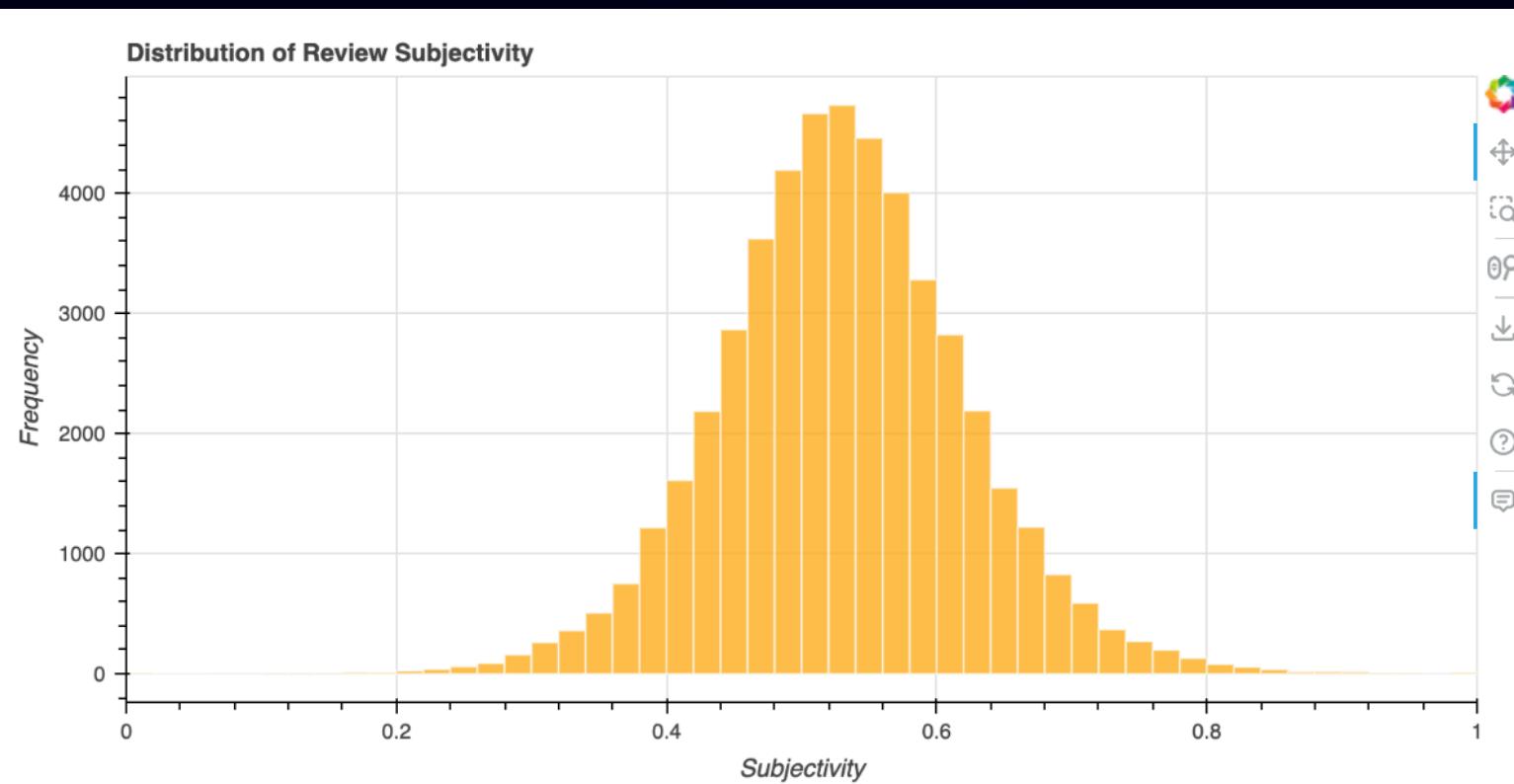
Neutral, detailed insights



Polarity and Subjectivity Insights

Polarity

Most reviews are moderately positive
(0 to 0.5 polarity)



Subjectivity

Most reviews are moderately subjective
(0.5 to 0.7 subjectivity)

Audiences prefer balanced, moderate feedback rather than extreme opinions



User reviews

4,672 reviews Hide spoilers + Review

Ratings Show all ▾ Sort by Featured ▾ ↓

★ 9/10

This is why we watch movies. >

Fight Club does exactly what movies should do, blow our minds, make us ask questions of ourselves and our society, and all the while thoroughly entertain us. Edward Norton and Brad Pitt are at their very best.

Helpful · 134 7

Lk1990 · Apr 30, 2020 · Permalink

★ 8/10

We should actually talk about the Fight Club >

Spoiler ▾

Helpful · 538 46

ArthvrPendragOn · Nov 17, 2019 · Permalink

Patterns in Extreme Reviews

Shorter Reviews

- 10-50 words
- Stronger emotions
- Highly positive or highly negative*

Longer Reviews

- >1000 words
- Neutral and detailed
- Less emotional*

Limitations & Future Work

The analysis provided useful insights into user sentiment, but some challenges and areas for improvement remain.

Limitations

- TextBlob struggles with sarcasm or complex tones
- Minimal preprocessing → leftover text noise
- No reviewer context (e.g., helpful votes, user profiles, genre)

Future work

- Use transformer models (like BERT) for deeper sentiment detection
- Add metadata to uncover richer patterns
- Expand to other review sites for cross-platform analysis

